

University of Dundee

## Nuclear DNA Replication in Trypanosomatids

da Silva, Marcelo S.; Pavani, Raphael S.; Damasceno, Jeziel D.; Marques, Catarina A.; McCulloch, Richard; Tosi, Luiz Ricardo Orsini

*Published in:*  
Trends in Parasitology

*DOI:*  
[10.1016/j.pt.2017.08.002](https://doi.org/10.1016/j.pt.2017.08.002)

*Publication date:*  
2017

*Licence:*  
CC BY

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

### *Citation for published version (APA):*

da Silva, M. S., Pavani, R. S., Damasceno, J. D., Marques, C. A., McCulloch, R., Tosi, L. R. O., & Elias, M. C. (2017). Nuclear DNA Replication in Trypanosomatids: There Are No Easy Methods for Solving Difficult Problems. *Trends in Parasitology*, 33(11), 858-874. <https://doi.org/10.1016/j.pt.2017.08.002>

### General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Review

## Nuclear DNA Replication in Trypanosomatids: There Are No Easy Methods for Solving Difficult Problems

Marcelo S. da Silva,<sup>1</sup> Raphael S. Pavani,<sup>1</sup>  
 Jeziel D. Damasceno,<sup>2</sup> Catarina A. Marques,<sup>3</sup>  
 Richard McCulloch,<sup>4</sup> Luiz Ricardo Orsini Tosi,<sup>2</sup> and  
 Maria Carolina Elias<sup>1,\*</sup>

**In trypanosomatids, etiological agents of devastating diseases, replication is robust and finely controlled to maintain genome stability and function in stressful environments. However, these parasites encode several replication protein components and complexes that show potentially variant composition compared with model eukaryotes. This review focuses on the advances made in recent years regarding the differences and peculiarities of the replication machinery in trypanosomatids, including how such divergence might affect DNA replication dynamics and the replication stress response. Comparing the DNA replication machinery and processes of parasites and their hosts may provide a foundation for the identification of targets that can be used in the development of chemotherapies to assist in the eradication of diseases caused by these pathogens.**

### Trypanosomatids, a Group of Eukaryotes with Peculiar Features

Trypanosomatids are a group of parasitic single-celled eukaryotes within the order Kinetoplastea. Amongst the trypanosomatids are human pathogens of paramount medical importance, such as *Leishmania* spp. (etiological agent of distinct forms of leishmaniasis), *Trypanosoma cruzi* (etiological agent of Chagas' disease), and *Trypanosoma brucei* (etiological agent of African sleeping sickness). Altogether, these parasites are responsible for more than 50 000 deaths annually [1]. Trypanosomatids present a heteroxenous life cycle (i.e., they require more than one host to complete their life cycle), varying between replicative (usually noninfective) and nonreplicative (infective) forms, which makes one wonder if genome replication and infection could be mutually exclusive events. They diverged from other eukaryotes around 200–500 million years ago (MYA) [2–4], which comprises the period between the emergence of arthropods and mammals (Figure 1). This timing suggests that trypanosomatids diverged as a result of new niches provided by the metazoans, which allowed trypanosomatids to coevolve with them and led to the emergence of parasitic and symbiotic relationships [3–5]. Associated with this evolution, trypanosomatids present several unique characteristics amongst eukaryotes, including the near universal use of multigenic RNA polymerase (Pol) II transcription and, in *T. brucei*, adaptation of RNA Pol I to transcribe genes encoding surface proteins. These peculiar features might be explained by the fact that these organisms belong to the phylum Euglenozoa, maintaining the 'primitive' characteristics of this higher order taxon, or

### Trends

In trypanosomatids, DNA replication is tightly controlled by protein complexes that diverge from those of model eukaryotes.

There is no consensus for the number of replication origins used by trypanosomatids; how their replication dynamics compares with that of model organisms is the subject of debate.

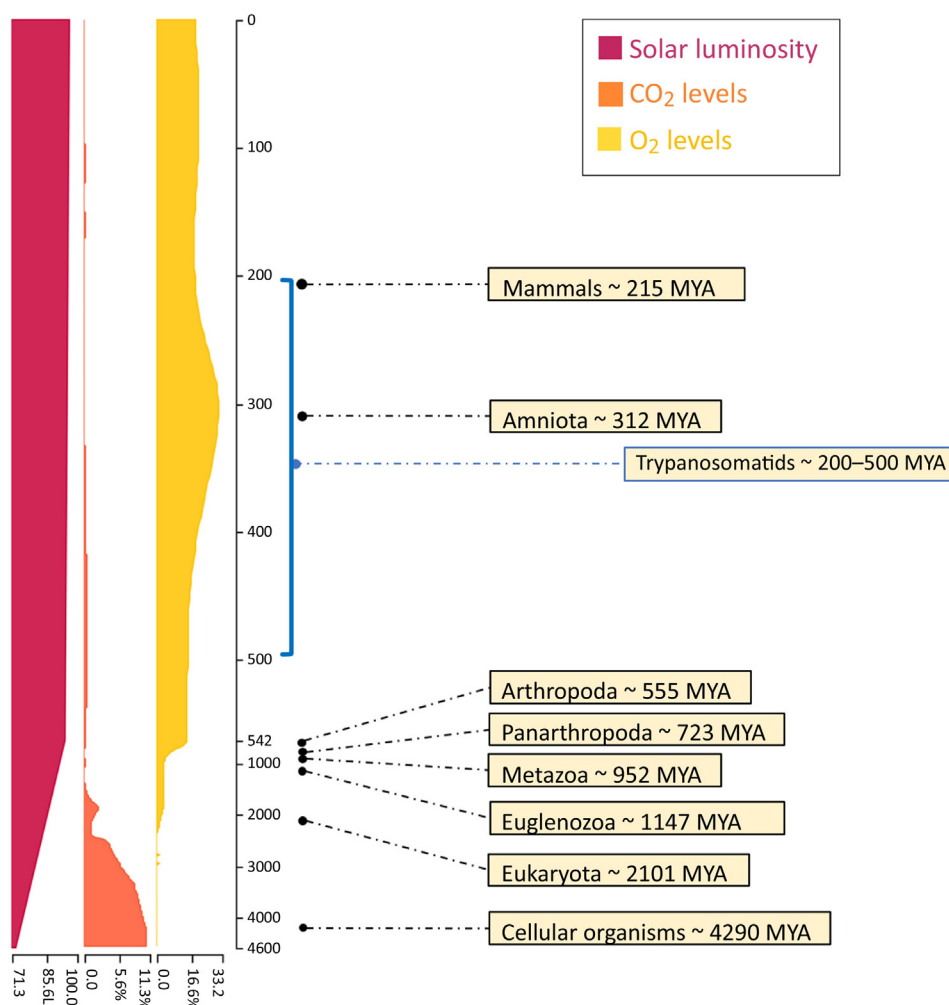
The DNA replication rate in trypanosomatids is similar to, but slightly higher than, that of model eukaryotes, which may be related to chromatin structure and function.

Recent data suggest that the origin recognition complex in trypanosomatids closely resembles the multisubunit eukaryotic model.

The absence of fundamental replication-associated proteins in trypanosomatids suggests that new signaling pathways may be present in these parasites to direct DNA replication and the replicative stress response.

<sup>1</sup>Laboratório Especial de Ciclo Celular (LECC), Center of Toxins, Immune Response and Cell Signaling (CeTICS), Butantan Institute, São Paulo, SP, Brazil

<sup>2</sup>Departamento de Biologia Celular e Molecular e Bioagentes Patogênicos,



## Trends in Parasitology

**Figure 1. Evolutionary Timeline of Eukaryotes.** Trypanosomatids diverged around 200–500 million years ago (MYA), probably due to the emergence of new niches precipitated by the appearance of the metazoans, which allowed coevolution of trypanosomatids with both vertebrate and arthropod hosts. Adapted and modified with permission from [www.timetree.org](http://www.timetree.org).

by genome streamlining, since the transition from a free-living to a parasitic lifestyle resulted in the loss of many protein-coding genes in these organisms [6,7]. Whether the unusual composition and structure of at least some of the proteins involved in DNA replication in trypanosomatids, as well as the unique dynamics of the reaction, reflect ancient peculiarities or adaptive events remains unclear.

In general, the dynamics and mechanics of DNA replication are conserved among living organisms and fine-tuned to cellular growth. In eukaryotes, at the end of mitosis and during the G1 phase, protein complexes are assembled onto sites in the DNA genome, establishing these locations as potential regions for the onset of new DNA synthesis (these regions are called replication origins). When the S phase is reached, some of these established origins will be activated, at which time DNA replication proceeds by the recruitment of further protein complexes. Understanding the processes involved in trypanosomatid DNA replication could not merely provide an evolutionary perspective on this central cellular reaction but may help to

Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (USP), Ribeirão Preto, SP, Brazil

<sup>3</sup>Division of Biological Chemistry and Drug Discovery, School of Life Sciences, Dow Street, University of Dundee, Dundee, UK

<sup>4</sup>The Wellcome Centre for Molecular Parasitology, Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, UK

\*Correspondence:

[carolina.eliassabbaga@butantan.gov.br](mailto:carolina.eliassabbaga@butantan.gov.br)  
(M.C. Elias).

explain the striking ability of these parasites to gain, lose, or rearrange DNA, allowing for a better adaptation to the environment [8,9]. Knowledge about genome maintenance pathways, and in particular the process of nuclear DNA replication, in any trypanosomatids, is limited compared to that of model eukaryotes. However, several studies have emerged recently that have disclosed many unusual features of the replication machinery, the control of re-replication and the replication stress responses, elevating trypanosomatid parasites to the forefront of understanding of nuclear DNA replication and transmission amongst protozoans, which provide much of the diversity of the eukaryotic domain of life. In this review, we focus on the main advances in recent years regarding the identification of replication origins, the structure of the protein complexes involved in nuclear DNA replication, and mechanisms to avoid re-replication and cope with replicative stress in the main pathogenic trypanosomatids. Moreover, by comparing these processes with what is known in model eukaryotes, we hope to leave clues in a trail of discovery that will ask if important components or controls of nuclear DNA replication could emerge as a target for the development of new antitrypanosomatid therapies, helping to eradicate the diseases caused by these parasites.

### DNA Replication Origins

DNA replication origins are strictly defined as sites in the DNA genome that are bound by specialized initiator proteins in order that the wider replication machinery can be loaded and the onset of genomic DNA synthesis can begin. In general, bacteria typically have a single replication origin per genome, archaea can have a single origin or can use a few origins, whereas each model eukaryote genome is replicated from hundreds or thousands of origins. Currently, there is little clear consensus for the number of replication origins used by trypanosomatids, meaning that how their replication dynamics compare with prokaryotes and eukaryotes is the subject of debate.

The first study in trypanosomatids to identify and count the number of DNA replication origins was a genome-wide analysis performed in *T. brucei* using a technique called MFaseq (Box 1) [10]. This assay showed that *T. brucei*, like all previously characterized eukaryotes, presents multiple origins per chromosome, with peak amplitude variation, as obtained by MFaseq, suggesting differing timing or frequency of activation. However, *T. brucei* origins are more widely spaced than in other eukaryotes: one origin for each 260 kb, compared with budding yeast, where there is one per every 46 kb, and mammalian cells, where there is one every 25 to 130 kb [10]. Nonetheless, correlating MFaseq peak location and the binding sites of a replication-initiating factor (see below) showed that *T. brucei* licenses more origins than are activated during the S phase. Though origins and initiator-binding sites all localize to the ends of the multigene transcription units, no consensus sequence for origins was found [10].

Remarkably, when MFaseq was applied to *Leishmania major* and *Leishmania mexicana*, only a single replication initiation site per chromosome was found [11], an unprecedented observation in eukaryotes. Considering that trypanosomatids perform polycistronic transcription and there might be transcription during replication, the process of origin activation in these parasites might need to strike an equilibrium between the activation of as few origins as possible to limit replication–transcription conflicts, and the activation of sufficient origins to allow replication of the entire genome. In this scenario, it appears that different solutions have been reached by each parasite. However, two other techniques may indicate that there are more origins activated in *T. brucei* and *Leishmania* spp. than are detected by MFaseq. One of these techniques is called DNA combing (Box 1), which was used to visualize replication in a specific region of *T. brucei* chromosome 1, revealing initiation at a single non-MFaseq mapped initiator binding site, and therefore origin, after hydroxyurea treatment [12]. DNA combing was also applied to undefined DNA molecules in *Leishmania* spp. and *T. brucei*, revealing more than a single initiation site in *Leishmania* spp and less widely spaced initiation sites in *T. brucei*, but

### Glossary

**Chromatin:** complexes of macromolecules on the nuclear genome of eukaryotic cells.

**Constitutive origins:** DNA replication origins that are activated in all cells of a population.

**Decatenating activities:** the process of separating topologically-linked circular chromosomes.

**Dormant origins:** DNA replication origins that are not fired during a normal cell cycle but are activated in the presence of DNA damage/replication stress.

**Flexible origins:** DNA replication origins whose use varies from cell to cell in an apparently stochastic manner.

**GINs:** an acronym from the first letters of the Japanese numbers 5-1-2-3 (go-ichi-ni-san) in a reference to the proteins Sld5, Psf1, Psf2, and Psf3.

**Kinetoplast:** a unique mitochondrial DNA structure organized in a giant network of interlocked rings (also called kDNA), found exclusively in flagellated protists belonging to the group Kinetoplastea.

**Okazaki fragments:** short, newly synthesized DNA fragments (with an RNA primer on the 5' end) generated on the lagging strand during DNA replication.

**Pol  $\alpha$ -primase complex:** DNA polymerase alpha is an enzyme complex from eukaryotes composed of four subunits: POLA1 (catalytic subunit), POLA2 (regulatory subunit), PRIM1 (small primase subunit), and PRIM2 (large primase subunit).

**Replisome:** complex machinery composed of several proteins, including MCM<sub>2-7</sub>, CDC45, GINS, RFC, PCNA, RPA, topoisomerases, pol  $\alpha$ -primase, pol  $\delta$ , pol  $\epsilon$ , FEN-1, and DNA ligase. Its main function is to carry out replication of DNA.

**Sumoylation:** a post-translational protein modification by addition of SUMOs (small ubiquitin-like modifiers).

**Telomerase:** a reverse transcriptase enzyme responsible for elongating telomeres by addition of a species-dependent telomere repeat sequence to the 3' overhang.

**Telomeres:** a region at the end of eukaryotic chromosomes, composed of repetitive sequences and associated proteins, whose main function is to protect the chromosome from the action of

without correlation to initiator-binding sites or wider chromosome features [13]. The other approach used, called SNSseq (Box 1), revealed more than 5000 sites of replication initiation throughout the *Leishmania* chromosomes, with most sites dispersed throughout the multigene transcription units [14].

The DNA combing and SNSseq studies may suggest that there are more active origins in *T. brucei* and *Leishmania* spp. than those detected by MFaseq, which provides an indication of the frequency with which an origin is activated in a cell population [10,11], meaning that MFaseq might detect mainly **constitutive origins** (see Glossary). Indeed, some authors have suggested that MFaseq peaks actually represent sites of pronounced concentration of many origins [14,15]. In contrast, DNA combing and SNSseq techniques may not reflect the frequency of origin activation, meaning that these techniques might also identify **flexible, possibly dormant origins**. However, it is important to stress that only MFaseq has, to date, been correlated with initiator binding, so it is also possible that unconventional, origin-independent replication initiation events are also being detected by the other two techniques. In conclusion, we propose that *Leishmania* and *T. brucei* present constitutive origins that are fired in every cell cycle, with further flexible sites of replication initiation being fired stochastically. Thus, the mechanisms that determine the latter sites of activation need to be determined, including whether the same strategies to complete DNA replication after potential replication-transcription conflicts are used in *T. brucei* and *Leishmania*. Indeed, the characteristics of constitutive origins, such as histone modifications that might recruit the initiator protein

exonucleases and from fusion with other chromosomes.

#### Variant surface glycoprotein

**genes:** genes which encode proteins that pack the cell surface of *T. brucei* and allow this parasite to evade the host's immune system by antigenic variation.

### Box 1. Techniques Used to Monitor DNA Replication Origins in Trypanosomatids

#### DNA Combing

This technique is used to produce an array of uniformly stretched DNA molecules, allowing the investigation of DNA replication on single molecules. It requires two consecutive pulses of thymidine analogs (usually IdU and CldU) in an asynchronous culture of cells. Usually, the cells with IdU-CldU incorporated are trapped in agarose plugs and the DNA is isolated by treatment with proteinase K. For analysis of a specific fragment of the genome, the plug containing DNA is subjected to pulsed-field gel electrophoresis (PFGE). Part of this PFGE is submitted to Southern blotting, which will detect a fragment of interest through hybridization with specific probes. The fragment of interest is then recovered from the other part of PFGE, treated, and stretched (combed) on slides. Alternatively, for an analysis of the whole genome, the plug containing DNA treated with proteinase K can be directly stretched (combed) on the slides. IdU, CldU, and DNA are detected by indirect immunofluorescence using specific antibodies. Different signal patterns for DNA replication origins can be observed in the slide analysis, allowing visualization of replication fork direction, including initiation and termination regions (Figure 1).

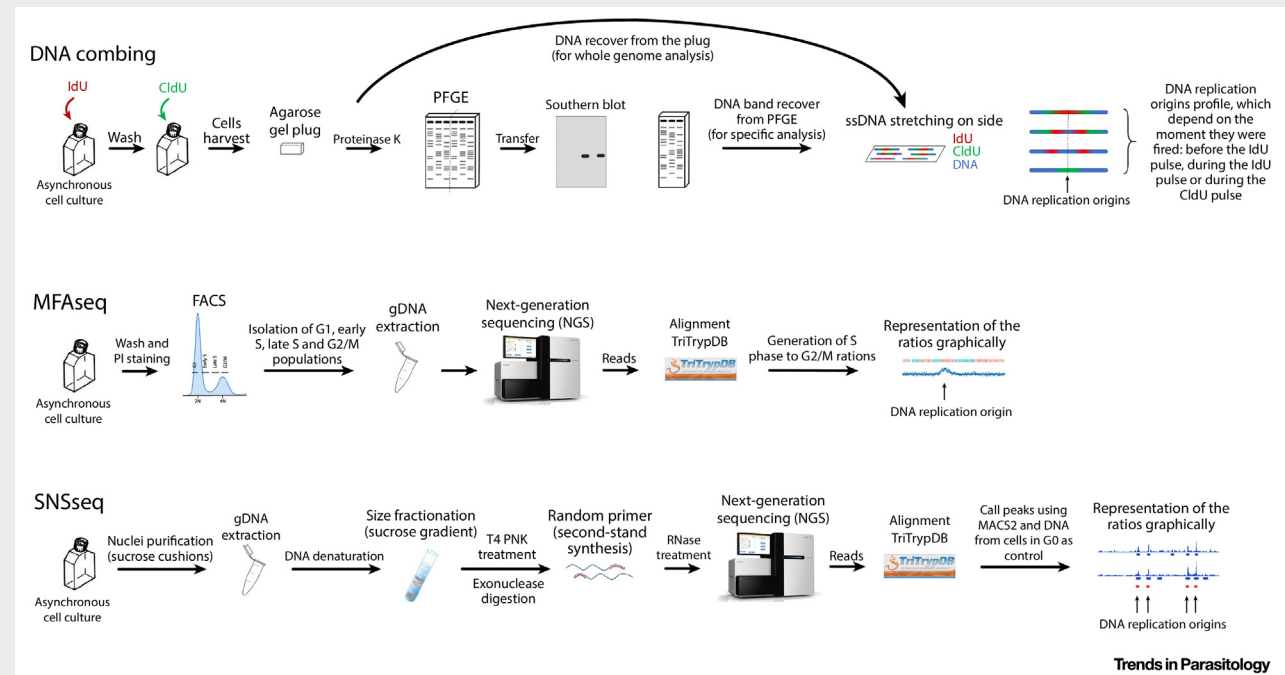
#### MFaseq

'Marker frequency analysis coupled with deep sequencing', also termed Sort-seq in yeast and archaea, is a population-based analysis in which DNA read depth across the genome of replicating cells is compared relative to that in nonreplicating cells. This results in a landscape of the replication profile across the genome, in which 'peaks' represent replicating regions of the chromosome (regions where DNA replication origins are landed to be present), while 'valleys' (representing the convergence of two replication forks) suggest replication termination zones. Moreover, the peak height and width can also provide information on origin usage and replication speed. Briefly, an asynchronous culture of cells is fixed and stained with a DNA dye (e.g., propidium iodide) and sorted by fluorescence-activated cell sorting (FACS) into populations of the different cell cycle stages. Sorting the S phase into two populations, early and late, is optional but allows the analysis of the replication profile at two different stages, and the mapping of early and late replicating regions of the genome. Once the populations have been isolated, genomic DNA is extracted, prepared for next-generation sequencing (NGS), and sequenced. The resulting reads are then aligned to the reference genome (which can be retrieved from TriTrypDB.org) and the frequency of reads (coverage) is assessed per base pair, which in this analysis acts as a 'marker'. For each population, the genome is then 'fragmented' into bins (analysis to date used 1–2.5 kbp bins) and the median of the coverage per bin is calculated (for each sample, the median of read depth per bin is compared to the general genome-wide average of reads of that sample). Coverage is then used to compute the ratio between the replicating populations (S phase) and the nonreplicating ones (G1 or G2/M phases) per bin, which is then plotted and represented graphically across the genome (Figure 1).

#### SNSseq

'Small nascent strand purification coupled with deep sequencing' is a highly sensitive, population-based analysis for sequencing initiation-proximal nascent DNA strands that are linked to RNA primers (Figure 1) and is widely used in studies of replication in metazoans; it is capable of detecting initiation sites used by only a few cells in the population. Simplistically, nuclei are purified from an asynchronous culture of cells using sucrose cushions, and the genomic DNA is extracted. The DNA is then heat-denatured and size-fractionated using a range of sucrose gradients. Fragments ranging from 300 to 1500 nucleotides are then purified and subjected to rounds of T4 polynucleotide kinase (PNK) phosphorylation, followed by digestion with  $\lambda$ -exonuclease. The treated DNA is then subjected to random-primed second-strand synthesis and next digested with RNase. The resulting DNA is then processed to generate next-generation sequencing (NGS) libraries, and sequenced. The

reads obtained are then aligned to the reference genome, and peaks are called using the callpeak function of MACS2 (using processed DNA from cells collected from a cell culture in stationary phase as a control). The peak-profile can then be plotted and represented graphically across the genome (Figure 1).



**Figure 1. The Main Steps of the Techniques Used to Monitor DNA Replication in Trypanosomatids.** DNA combing, MFaseq, and SNSseq. IdU, 5-iodo-2'-deoxyuridine; CldU, 5-chloro-2'-deoxyuridine; PFGE, pulsed-field gel electrophoresis; ssDNA, single-stranded DNA; FACS, fluorescence-activated cell sorting; gDNA, genomic DNA; NGS, next-generation sequencing; T4 PNK, polynucleotide kinase from T4 bacteriophage. Of note, these key steps were described in previous studies [12–14,21,24].

complexes or colocalization with centromeres, increasing the probability of these origins being fired, need to be addressed.

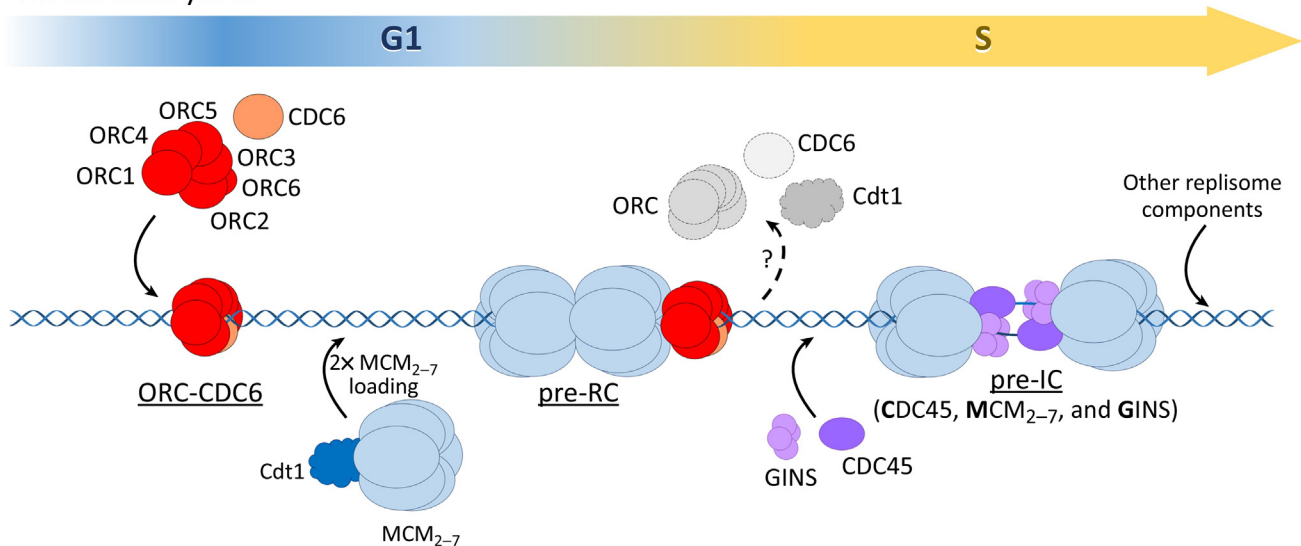
### Pre-Replication (Pre-RC) and Pre-Initiation (Pre-IC) Complexes

The establishment of a DNA replication origin occurs by the ordered recruitment of protein complexes (Figure 2), leading to the assembly of the replicative machinery. In model eukaryotes, DNA replication origins are demarcated by the binding of a conserved six-subunit initiator protein complex called the origin recognition complex (ORC), whose activity is modulated by the binding of another, related protein termed cell division cycle subunit 6 (CDC6). Five of the six ORC subunits (ORCs<sub>1–5</sub>) and CDC6 are AAA+ family ATPases and possess C-terminal winged helix (WH) DNA-binding domains. The ORC structure resembles an open two-layered ring with AAA+ ATPase subunits in one layer and WH domains in the other [16,17]. ORC6 is the only subunit that does not conform to the ORC AAA + WH structure, and its location within the complex remains somewhat unclear, though it appears to be adjacent to ORCs<sub>2–3</sub> [18,19]. Binding of CDC6 between ORC1 and ORC2 appears to close the ring and allows ORC–CDC6 to recruit the replicative helicase called the mini-chromosome maintenance (MCM) complex, which is also composed of six AAA+ ATPase subunits (MCM<sub>2–7</sub>). Productive and stable interaction between ORC–CDC6 and MCM<sub>2–7</sub> requires a further mediator, termed CDC-dependent transcript 1 (Cdt1), which can bind stably to MCM6 [18,19]. Though structural studies show the MCM<sub>2–7</sub> helicase is loaded as a double hexamer, and the precise mechanics of loading the two hexamers remain unclear, CDC6 also plays a central role in this reaction [20].

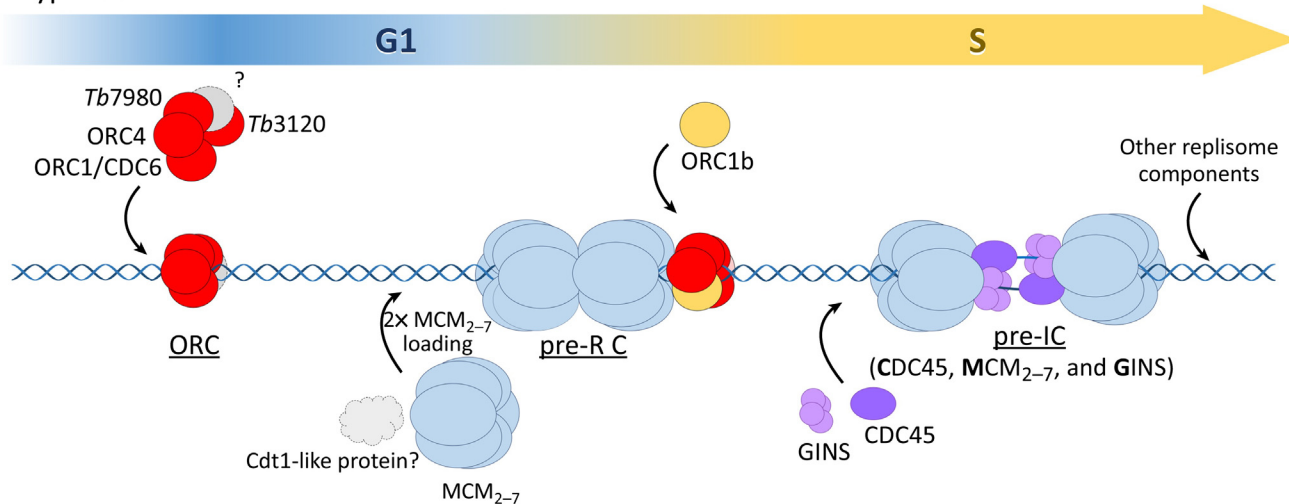


In total, ORC–CDC6–Cdt1–MCM<sub>2–7</sub> (termed the prereplication complex, or pre-RC) assembles onto origins during the G1 phase. Activation of the pre-RC occurs in the S phase and involves interaction of each MCM<sub>2–7</sub> hexamer with a single protein termed cell division cycle subunit 45 (CDC45) and a four-subunit complex called **GIN5**. The resulting CDC45–MCM<sub>2–7</sub>–GIN5 (CMG)

### Model eukaryotes



### Trypanosomatids

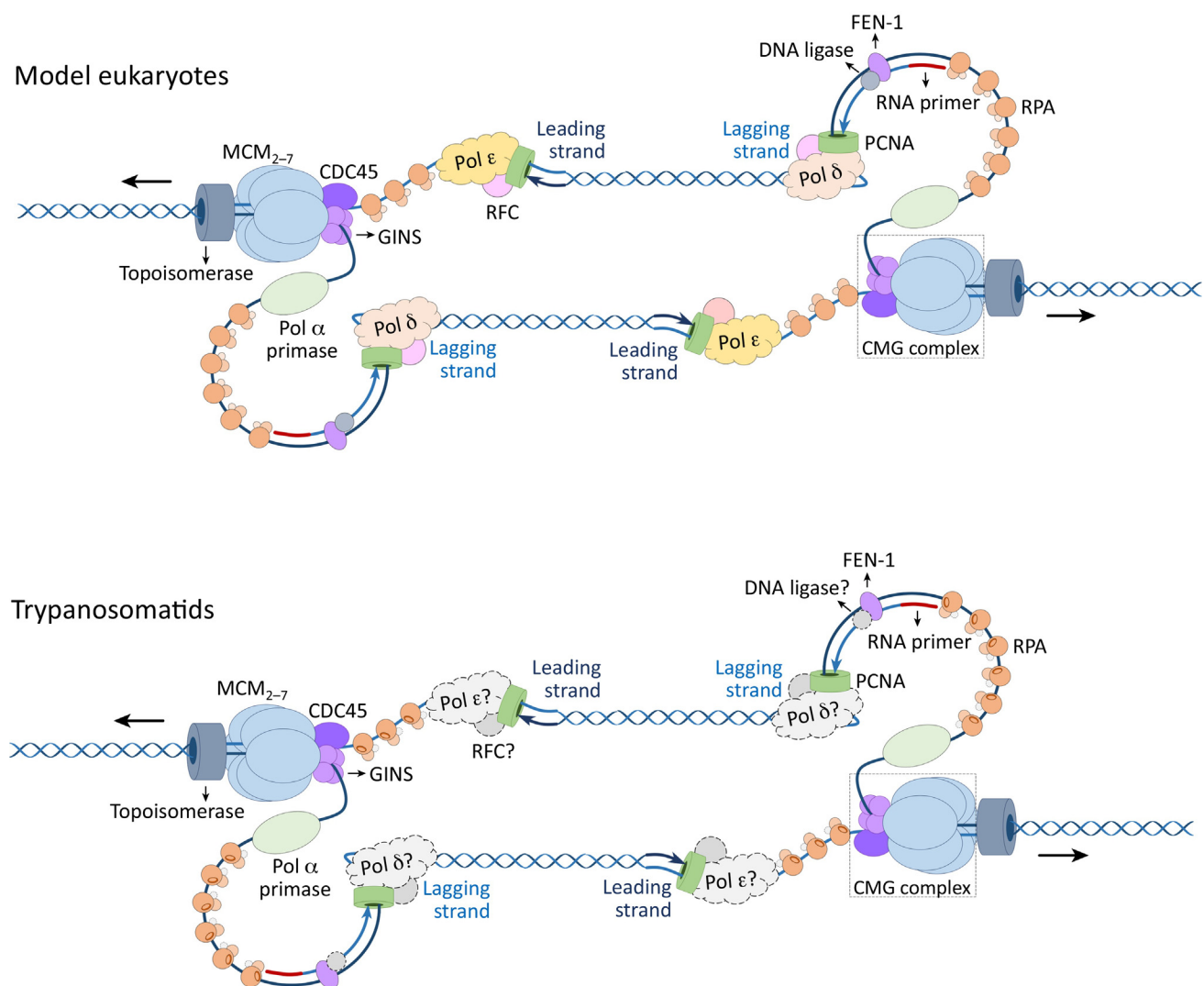


Trends in Parasitology

**Figure 2. The Replication Initiation Machinery of Model Eukaryotes and Trypanosomatids.** In model eukaryotes (top panel), from late mitosis to the end of G1 phase, the six-subunit ORC is recruited to all potential origins in the genome. Here, ORC interacts with CDC6. Next, the MCM<sub>2–7</sub> helicase is recruited and loaded onto the origin by Cdt1. This forms the pre-RC, and renders the origins of replication 'licensed'. At the onset of the S phase, CDC6 and Cdt1 are removed from the pre-RC, CDC45 and GIN5 are recruited to the origin, forming the pre-IC. Together, CDC45, the MCM<sub>2–7</sub>, and the GIN5 complex form the CMG complex, which is the active replicative helicase that unwinds the origin DNA, allowing the further assembly of the replicative fork components. In trypanosomatids (bottom panel), a divergent ORC-like complex is present, comprising ORC1/CDC6, ORC4, Tb3120, and Tb7980. It is not clear if other subunits remain to be identified. It is assumed that MCM<sub>2–7</sub> is loaded, as in other eukaryotes, prior to the S phase, but how this is catalyzed is unclear, as no Cdt1 orthologue has been identified, and a clear orthologue of CDC6 remains undetermined. It remains possible that MCM is not loaded until the S phase (not shown). At the end of G1 and the start of the S phase, until late G2, an ORC1 orthologue, ORC1b, is expressed. CDC45 and the GIN5 complex are, most likely, then recruited to the origin, and the steps downstream are believed to take place in a similar way to model eukaryotes. ORC, origin recognition complex; CDC6, cell division cycle subunit 6; MCM, mini chromosome maintenance; Cdt1, CDC-dependent transcript 1; pre-RC, pre-replication complex; CDC45, cell division cycle subunit 45; GIN5, 'go-ichi-ni-san' in reference to the proteins Sld5, Psf1, Psf2, and Psf3; pre-IC, pre-initiation complex; CMG, complex composed of CDC45, MCM<sub>2–7</sub>, and GIN5.

complex, also termed the pre-initiation complex (pre-IC), possesses helicase activity, and its formation is associated with MCM phosphorylation by the CDC7-Dbf4 kinase complex. Once formed, the CMG complex is able to recruit the other **replisome** components to melted DNA at the origin, a reaction that involves further factors, including MCM10 and a RecQ helicase (not shown in Figures 2 and 3).

Recent studies have begun to reveal pronounced divergence in the origin-binding initiator of trypanosomatids, when compared with model eukaryotes. In *T. brucei*, MCM<sub>2-7</sub>, CDC45, and GINS are conserved and essential (Box 2). Although functional analyses have not been extended beyond *T. brucei*, the conservation of the various components in *T. cruzi* and *Leishmania* suggest that the CMG machinery in trypanosomatids is very similar that in to



Trends in Parasitology

**Figure 3. Replication Fork Progression in Trypanosomatids.** All proteins belonging to the replication fork (topoisomerases, MCM<sub>2-7</sub>, CDC45, GINS, RPA, PCNA, and Pol α-primase) presented in color have been described in trypanosomatids. Genes encoding homologues of DNA polymerases δ and ε, RFC, and DNA ligase I (presented in gray, with a dashed border) are present in the sequenced genomes of trypanosomatid parasites, but their role in replication fork progression has not yet been characterized. MCM, mini chromosome maintenance; CDC45, cell division cycle subunit 45; GINS, 'go-ichi-ni-san' in reference to the proteins Sld5, Psf1, Psf2, and Psf3; RPA, replication protein A; PCNA, proliferating cell nuclear antigen; RFC, replication factor C; FEN-1, flap endonuclease 1.



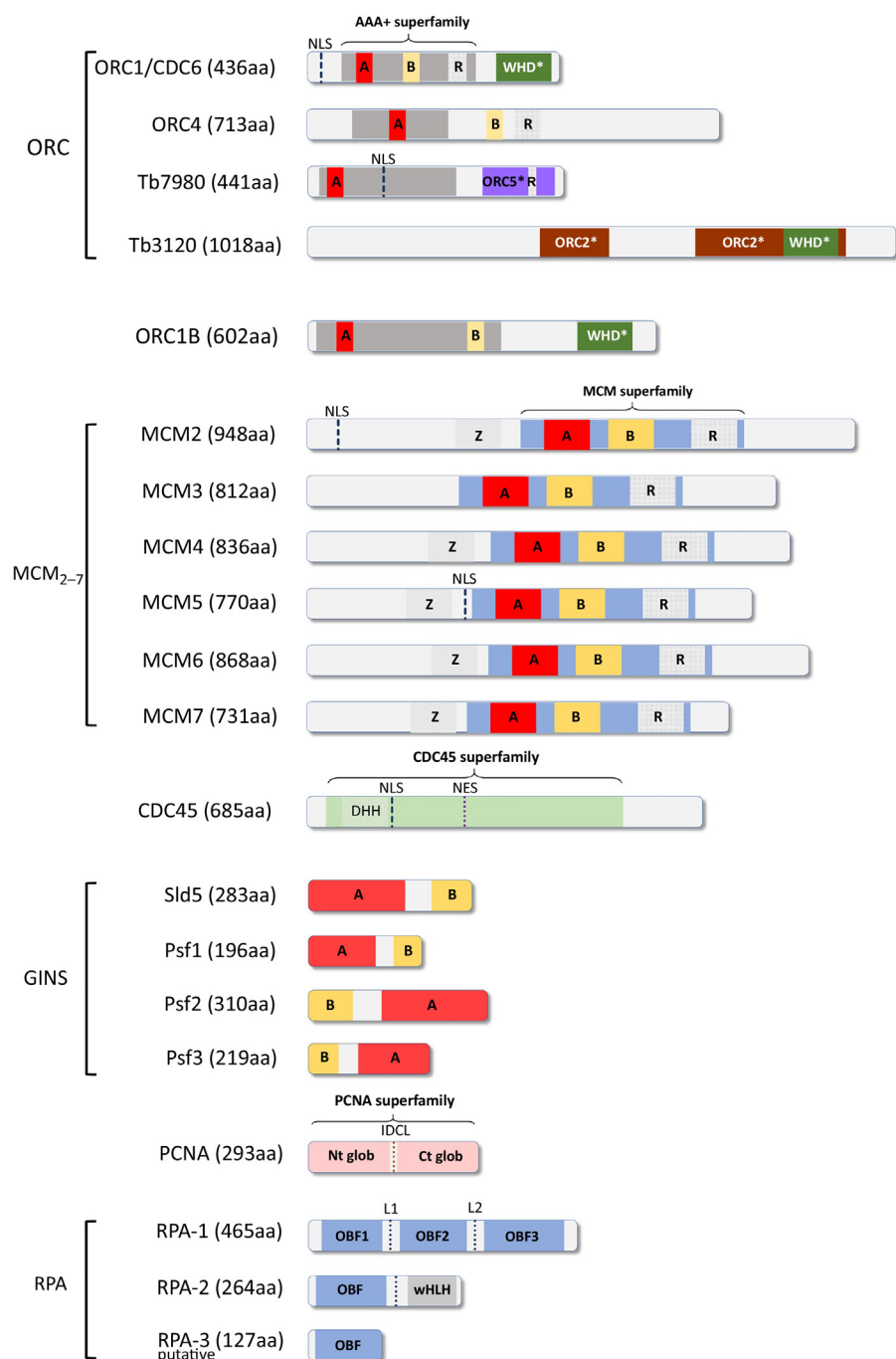
model eukaryotes [21,22]. In contrast, trypanosomatid ORC has proved much more difficult to characterise. Initially, trypanosomatid genome searches identified only a single ORC-like factor (named ORC1/CDC6), homologous to ORC1 and CDC6 and capable of complementation of yeast CDC6 mutants [23]. Given the lack of identifiable orthologues of other ORC subunits, it was considered possible that ORC activity resides in a single factor, as is seen in members of the Archaea. Consistent with this hypothesis, no homologue of Cdt1 has yet been identified in any kinetoplastid. However, subsequent work revealed that *T. brucei*, and most likely all kinetoplastids, does in fact possess an ORC, albeit one that appears highly dissimilar to the conserved six-subunit ORC–CDC6 initiator seen in model eukaryotes (Figure 2). The first evidence of a diverged ORC came from the identification of a weakly conserved ORC1-like factor, which was named ORC1B and shown to interact with ORC1/CDC6 and MCM3 in *T. brucei* [22]. Later, three more ORC1/CDC6-interacting factors were identified in *T. brucei*: a highly divergent ORC4-like subunit, and Tb7980 and Tb3120, two factors with very limited primary sequence homology with ORC subunits [21]. Subsequently, RNAi revealed that loss of ORC1/CDC6, ORC1B, ORC4, or Tb3120 impedes DNA replication and leads to comparable growth and cell cycle defects [24]. Though loss of Tb7980 results in proliferation defects [21], clear evidence of a role in DNA replication has not yet been determined, but is likely. The strongest evidence for an ORC in *T. brucei* is found in the demonstration that ORC1/CDC6 and ORC4 are present in a high-molecular-weight complex (~530 to 1011 kDa) that also seems to include MCM3, though whether all putative ORC and MCM subunits are also present is currently unknown [24]. Of note, localization studies suggest that ORC1B is unlikely to be a stable component of *T. brucei* ORC [24]. Nonetheless, ORC1/CDC6-binding sites have been mapped in the *T. brucei* genome and shown to colocalize with sites of replication initiation, clearly showing that at least one ORC component dictates constitutive origin function [10]. Around 60% of ORC1/CDC6 binding sites are found in subtelomeric chromosome regions containing **variant surface glycoprotein** (VSG) genes, where replication initiation has not been mapped. Whether this dense binding might relate to VSG expression control is currently unclear [25,26].

## Box 2. Structural Analysis of the Proteins Involved in Replication

Many homologues of replication proteins annotated in TriTrypDB possess structural differences compared to model eukaryotes.

- (a) ORC-related proteins (ORC1/CDC6, ORC4, Tb7980, Tb3120, and ORC1b): reviewed in [21].
- (b) CDC45: shares 20–5% identity with CDC45 from yeast and mammals. Despite the low identity, the conserved DHH domain is present [22]. Nuclear localization and export signals were also found in this counterpart [22].
- (c) MCM<sub>2–7</sub>: all possess the conserved ATPase domain found in model eukaryotes, containing Walker A, Walker B, and arginine-finger motifs [22,27]. Moreover, a zinc-finger motif (important for double hexamer formation) was also found in the N-terminal region in five out of six MCMs, suggesting that this arrangement can be formed [22].
- (d) GINS: organized by the arrangement of two conserved domains, A (rich in  $\alpha$  helices) and B (rich in  $\beta$  strands). Sid5 and Psf1 contain a large A N-terminal domain and a small B C-terminal domain, while Psf2 and Psf3 possess these two domains in exactly the opposite order, which is the same domain organization described for other eukaryotes [22].
- (e) PCNA: forms a homotrimeric ring complex and possesses conserved motifs, binding sites for DNA, and known interacting proteins, but contains an insertion of more than 30 residues in its C-terminal portion in all trypanosomatids [37,99].
- (f) RPA: composed of OB-fold domains similar to those in other eukaryotes but presents some important structural differences. The major subunit (RPA-1) lacks the first OB-fold domain (called 70N or DBD-F), which is important for protein–protein interactions [29,44,45]. RPA-2 contains one OB-fold domain and one wHLH in its C-terminal region, and is very structurally related to the model RPA-2. Although a possible RPA-3 homologue that was predicted to contain an OB-fold domain was found at TriTrypDB, no reports about its structure and function are available [29].
- (g) RFC and nuclear topoisomerases: although homologues of these proteins have been found in trypanosomatid genomes, no structural studies have been performed to date [8].
- (h) CDC6 and Cdt1: homologs of these proteins were not found in trypanosomatid genomes.

A scheme showing the main domains and motifs for these proteins is represented in Figure 1.



Trends in Parasitology

**Figure I. The Essential Domains and Motifs of the Main Proteins Involved in DNA Replication in *Trypanosomatids*.** NLS, nuclear localization signal; NES, nuclear export signal; Z, zinc-finger motif; A and B, Walker A and Walker B motifs; R, arginine fingers motif; WHD\*, putative winged helix domain; ORC5\*, putative ORC5 domain with low significance; ORC2\*, putative ORC2 domain with low significance; DHH, (Asp-His-His) motif; Nt and Ct glob, globular domains found at amino and carboxyl terminal regions of the protein; IDCL, inter-domain connecting loop; OBF, oligonucleotide/oligosaccharide-binding fold domain; wHLH, winged-helix-loop-helix domain; L, linkers between OBF domains.

ORC1B analysis suggests that the regulation of *T. brucei* (and therefore trypanosomatid) DNA replication may also be diverged. In contrast to ORC1/CDC6, ORC4, Tb3120, and Tb7980, which all localize to the *T. brucei* nucleus throughout the cell cycle, ORC1B localization and/or expression is highly cell-cycle-dependent and can only be detected in the nucleus of *T. brucei* cells in the S or G2 phase [24]. Thus, despite the sequence homology between ORC1B and ORC1 and/or CDC6, the expression and/or localization dynamics of this ORC-like factor do not resemble any eukaryotic ORC subunit thus far described (Figure 2). The active recruitment of ORC1B to the *T. brucei* nucleus during the S phase may indicate that the factor provides a positive regulatory role in DNA replication, which would be unprecedented [24]. How ORC1B might exert such an effect awaits experimental investigation, but its potential connection with the diverged kinetoplastid ORC is intriguing. The available data suggest that ORC1B is not a static member of the putative ORC-like complex and has drastically different expression dynamics to CDC6. However, is it possible that ORC1B provides modified CDC6-like functions? Binding of CDC6 converts ORC into a ring, allowing MCM<sub>2-7</sub> recruitment in model eukaryotes. In this light, it is conceivable that the diverged kinetoplastid ORC needs ORC1B to 'complete' the complex, but this happens at the onset of the S phase and not prior to the S phase. If so, the replication initiation cascade might be radically different from model eukaryotes. For instance, ORC-MCM<sub>2-7</sub> may interact and localize to origins in an inactive pre-RC that is only activated by binding of ORC1B (Figure 2). Alternatively, the pre-RC might only form at the outset of the S phase, when ORC1B allows MCM<sub>2-7</sub> to interact; in other words, MCM<sub>2-7</sub> is not recruited to origins until the S phase. Both possibilities may be consistent with the potential absence of Cdt1. However, they do not readily explain why ORC1/CDC6 is able to provide CDC6-like functions in being able to complement *Saccharomyces cerevisiae* CDC6 temperature-sensitive mutants [23]. Indeed, ORC1B has been suggested to lack ATPase activity, which perhaps renders it an unlikely candidate to mediate remodeling of the ORC [22]. Nonetheless, if any of these scenarios are correct, they raise questions that are presented in the outstanding questions box.

### Replication Fork Progression

After the establishment of a DNA replication origin and the assembly of the pre-IC complex, firing of replication occurs and replication is initiated with the bidirectional progression of replication forks from all origins. This process is conserved in the eukaryotic domain and, as trypanosomatids encode most of the predicted replication-fork proteins [22,24,27–29], we can update the description for replication fork progression in these organisms (Figure 3).

To open the double-stranded DNA and allow access to the DNA polymerases, trypanosomatids assemble the CMG complex [22]. RNA–DNA primers are generated by the **Pol  $\alpha$ -primase complex**. Then, DNA polymerase delta (Pol  $\delta$ ) and epsilon (Pol  $\epsilon$ ) synthesize DNA on the lagging and leading strands, respectively. It should be noted, however, that, in trypanosomatids, the composition and activity of Pol  $\delta$ , Pol  $\epsilon$ , and Pol  $\alpha$ -primase complexes remain to be fully characterized, though RNAi of a Pol  $\alpha$ -primase in *T. brucei* results in S phase arrest [30]. To promote the processivity of the replicative polymerases, a DNA sliding clamp called proliferating cell nuclear antigen (PCNA) strengthens the interaction between the DNA template and the polymerases. In most eukaryotes, PCNA is recruited by replication factor C (RFC) [31–33], which has not been examined in trypanosomatids. In *Leishmania* spp. and *Trypanosoma cruzi*, PCNA is concentrated in the nuclear periphery during the S phase [34,35], which is in agreement with localization of replicating DNA in the nuclear periphery [1,36]. On the other hand, *T. brucei* PCNA shows a dispersed pattern throughout the cell cycle [37], suggesting that DNA replication is organized differently within the nucleus of these trypanosomatids. PCNA and RFC form a moving platform for Pol  $\delta$  and play a number of roles, such as increased processivity, binding to the primer terminus, and bridging to the other replication proteins [33,38]. Furthermore, RFC and PCNA act as a loading platform for the flap

endonuclease-1 (FEN-1) and DNA ligase I, which process **Okazaki fragments** [33,39]. FEN-1 in *T. cruzi* was recently shown to participate in DNA replication and repair [40], but DNA ligase I has not yet been characterized in any trypanosomatid.

An essential protein for the progression of the DNA replication fork is replication protein A (RPA), which binds to ssDNA exposed at the replication fork to stabilize and protect this portion against nucleases. RPA helps in the recruitment of Pol  $\alpha$  to replicate the region where RNA priming is required for initiation [41,42] and reduces nucleotide misincorporation performed by Pol  $\delta$ , Pol  $\epsilon$ , and Pol  $\alpha$  [41–43]. In trypanosomatids, the best-characterized component of RPA is RPA-1, which lacks a 70N domain that is responsible for the interaction with repair proteins and protein kinases in model organisms [29,44–47]. In *T. cruzi*, RPA-1 and RPA-2 participate in DNA replication, while in *Leishmania amazonensis* RPA-1 seems to participate in double-strand break (DSB) repair, since it colocalizes with Rad51, a protein that catalyses homologous recombination [29,44,45].

Topoisomerases, which participate in resolving the over- or under-winding of DNA, constitute another group of proteins essential during replication fork progression. In trypanosomatids, topoisomerase type I is associated with the nucleolus and nuclear **chromatin** [33,48], while topoisomerase type II is essential for nuclear and kinetoplast DNA replication. Topoisomerase II proteins from *T. cruzi* and *Leishmania donovani* have ATP-dependent and ATP-independent **decatenating activities** [49–51]. In *L. donovani*, topoisomerase II appears in the nucleus and the **kinetoplast**, while in *T. cruzi* it is found only in the nucleus [33,49–51]. In most trypanosomatids, topoisomerase II shares high sequence identity and functions almost exclusively as a mitochondrial enzyme. However, two nuclear topoisomerase II enzymes (TbTOP2 $\alpha$  and TbTOP2 $\beta$ ) were described in *T. brucei*. They share similarity with nuclear topoisomerases from other eukaryotes, but TbTOP2 $\alpha$  encodes an ATP-dependent topoisomerase, whereas the role played by TbTOP2 $\beta$  remains unclear [51].

Under normal conditions, the replication fork of trypanosomatids, like all eukaryotes, continues synthesizing DNA until it reaches the chromosome **telomeres**. Due to the inability of the Pol  $\alpha$ -primase complex to initiate the last round of Okazaki fragment synthesis efficiently [52], telomeres exhibit a protrusion called a 3' overhang, which acts as a substrate for the enzyme **telomerase** to elongate the telomeres [53]. Of note, trypanosomatids differ in the protein components that bind single- or double-stranded telomeric regions, while retaining the same telomere repeat nucleotide structure, (TTAGGG) $_n$ , found in most eukaryotes [33,54,55].

The overall conservation in the replisome complexes in trypanosomatids, compared with model eukaryotes, is evident when we compare the speed of the DNA replication fork. Recent studies were able to calculate the replication speed in the different trypanosomatids using DNA combing: 1.84 kb/min for *T. brucei* strain Lister 427 [13], 3.7 kb/min for *T. brucei* strain TREU927 [12], 2.48 kb/min for *L. mexicana* [13], 2.45 kb/min for *L. major* [13], and 2.37 kb/min for *L. donovani* [13]. These rates range from 1.8 to 3.7 kb/min, which is broadly similar to the replication rates found in yeast (1.6 kb/min) [56], mouse embryonic fibroblasts (1.16 kb/min) [13], and in various human cell lines ( $\sim$ 1–2 kb/min) [57,58]. The small potential fork rate increase, relative to other eukaryotes, and potential differences between trypanosomatids, may be related to chromatin structure and function, reflecting multigenic transcription and the observation that trypanosomatid histones are divergent from those found in yeast and vertebrates and may undergo specific modifications [8,59,60].

### Mechanisms That Prevent Replication

While DNA replication is fundamental for cell proliferation, blockage of additional replication during and after the S phase is imperative for genomic stability. Hence, the firing of origins within

replicated regions should be avoided during the cell cycle. The compartmentalization of origin licensing at the G1 phase and origin firing at S phase is a perfect strategy to inhibit re-replication, because the S phase cyclin-dependent kinase (CDK) can fire origins at the same time as controlling pre-RC components to limit re-initiation at origins that have already fired. This control includes the modulation of pre-RC component expression and subcellular localization, in addition to their ability to interact with DNA. In yeast, CDC6 and Cdt1 are phosphorylated by CDK during the transition to G1/S phase, and these modifications can trigger protein degradation [61–63]. Non-DNA-bound MCM<sub>2–7</sub> is exported to the cytoplasm after the S phase [64]. Additionally, phosphorylation of ORC2 and ORC6 by CDK [65,66], as well as the interaction of CDK with CDC6, inhibits the recruitment of Cdt1/MCM onto origins [67]. Finally, the inhibition of all MCM subunits by **sumoylation** prevents helicase activation, which negatively regulates replication [68]. In metazoans, ORC1 and CDC6 undergo proteolysis after phosphorylation by CDK, and CDC6 is exported to the cytoplasm after CDK phosphorylation [69]. Phosphorylated MCM<sub>2–7</sub> is not exported to the cytoplasm but has a reduced activity for binding origins [70]. Since Cdt1 has an essential role in recruiting MCM<sub>2–7</sub> onto origins, metazoan cells rely on geminin, a Cdt1 inhibitor, to restrict replication to the S phase [71].

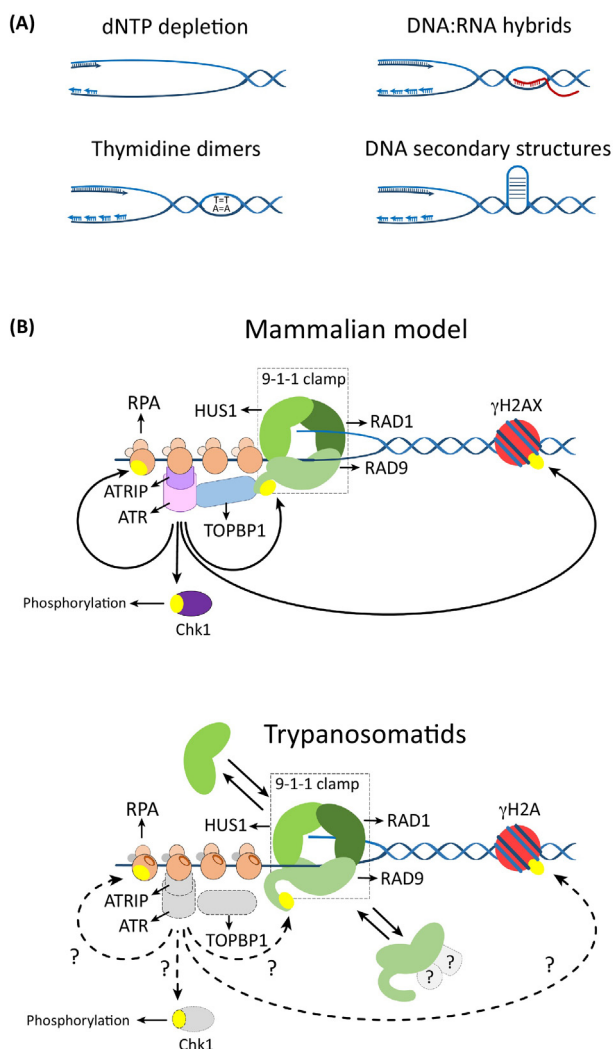
Re-replication events in the eukaryotic genome may lead to chromosome and gene copy number variation, which is a prevalent phenomenon in trypanosomatids, especially in the genus *Leishmania* [9]. However, other than ORC1B, no modulation of pre-RC component expression or subcellular localization has been described for any trypanosomatid. ORC1/CDC6 is present in the nuclear space in *L. major*, and remains bound to DNA throughout the cell cycle in *T. brucei* and *T. cruzi* [23,72]. In *T. brucei*, MCM<sub>2–7</sub> also remains in the nucleus throughout the cell cycle [22]. At this stage, due to the divergence of the pre-RC in trypanosomatids, it is not simple to infer how these organisms control DNA re-replication. Post-translational modifications, such as phosphorylation by CDKs, could act by avoiding helicase loading onto DNA after the S phase, but no protein kinase that controls DNA replication has been described. Moreover, as noted above, the unusual expression or localization of ORC1B, and the potential absence of Cdt1, has not been explored mechanistically in the context of replication control. Nonetheless, it has been shown that CDC45, a component of the pre-IC which activates MCM helicase activity together with GINS complex, is exported from the nucleus after the S phase in *T. brucei* [22]. Therefore, this may be a mechanism to prevent DNA re-replication, but how it occurs is unknown. DNA re-replication in other eukaryotes can lead to the generation of DNA DSBs and activation of the DNA damage checkpoint [73]. Considering that the potential mechanisms of gene amplification in *Leishmania* require factors that act in DSB repair [9], we can speculate on the existence of a possibly lax control to avoid the DNA re-replication and the pervasive structural variation exhibited by the genomes of *Leishmania* spp. It is possible that the same may apply to *T. cruzi*, where chromosome ploidy variation has been observed [74], though underlying mechanisms have not been explored. Chromosome ploidy changes in *T. brucei* appear limited to the VSG-rich subtelomeres, perhaps indicating a more rigid control of re-replication than in the other trypanosomatids [75].

In addition to replication prevention post-S phase, the DNA replication process is usually turned off in the infective forms of trypanosomatids. For instance, in *T. cruzi* trypomastigotes (infective form), ORC1/CDC6 is located in the nuclear space, but is not able to bind DNA. Moreover, MCM7 is not expressed in these parasite forms [27]. These data suggest that pre-RC components might be key factors in establishing replication arrest during the *T. cruzi* life cycle. Furthermore, *T. cruzi* probably uses different strategies to prevent DNA replication at the non-S phases of the replicative forms or in the infective ones, since ORC1/CDC6 is bound throughout the cell cycle in replicative forms but not bound to DNA in infective ones. The efficacy and details of these strategies, as well as their relation with the infection efficiency of *T. cruzi* and other trypanosomatids, requires further investigation.



## DNA Replication Stress

Replication stress is determined by events in which the DNA replication fork is slowed or halted, such as a decreased nucleotide pool, certain types of DNA lesions, RNA/DNA hybrids, or DNA secondary structures (Figure 4A). For some of these obstacles, there are translesion DNA Pols that help to overcome the DNA lesions [30]. However, in most cases the replicative helicase uncouples from the DNA Pol and continues to unwind the DNA duplex, inducing the accumulation of long stretches of single-stranded DNA (ssDNA) that are protected from degradation by association with RPA [76]. RPA-coated ssDNA mediates the recruitment of the ataxia telangiectasia Rad3-related (ATR) protein kinase and its binding partner ATR-interacting protein



**Figure 4. The Replication Stress Response in Mammals and Trypanosomatids.** (A) Obstacles that can slow or halt DNA replication leading to replication stress. These include dNTP depletion, DNA lesions (such as thymidine dimers), DNA secondary structures, and DNA:RNA hybrids. In all of these circumstances, the replicative helicase uncouples from the DNA polymerases, causing the accumulation of ssDNA and triggering replication stress-response. (B) The mammalian (top panel) and trypanosomatid (bottom panel) models for the replication stress response. Top panel. In this model, ATR kinase is recruited to RPA-coated ssDNA through its binding partner ATRIP. As an independent event, the 9-1-1 clamp is loaded at the ssDNA-dsDNA junction. TopBP1 is also recruited to RPA-ssDNA at the ssDNA-dsDNA junctions, interacts with the C-terminal tail of Rad9, and stabilizes the ATR kinase localization at the site of stress. ATR phosphorylates downstream factors, including H2A(X) and Chk1, which mediates cell cycle arrest and controls origin firing. Bottom panel. Most of the data for this pathway were observed in *Leishmania*. These parasites express a functional 9-1-1 homolog; the Rad9 subunit is found in alternative complexes, and Hus1 also exists as a monomer, suggesting a functional flexibility and compartmentalization of the trypanosomatid 9-1-1 clamp. Homologs of key elements of this pathway have not yet been characterized, such as the ATR-ATRIP complex, TOPBP1, or Chk1 (dashed gray molecules). dNTP, deoxynucleotide; ssDNA, single-stranded DNA; ATR, ataxia telangiectasia Rad3-related; RPA, replication protein A; ATRIP, ATR-interacting protein; Rad9, radiation sensitive subtype 9; Rad1, radiation sensitive subtype 1; Hus1, checkpoint protein HUS1; 9-1-1, complex composed of Rad9, Rad1, and Hus1; dsDNA, double-stranded DNA; TopBP1, DNA topoisomerase 2-binding protein 1; H2AX, histone variant; Chk1, checkpoint kinase 1.

Trends in Parasitology

(ATRIP) [77]. The ATR–ATRIP complex helps to stabilize and restart the collapsed fork and promotes the phosphorylation of substrates that include histone H2A(X) and checkpoint kinase 1 (Chk1) (Figure 4B) [78,79]. During these reactions, the ssDNA/dsDNA junction in these structures mediates the loading of a protein complex composed of Rad9, Rad1, and Hus1 (also called 9-1-1 clamp) [80]. The heterotrimeric 9-1-1 complex facilitates the recruitment of DNA topoisomerase 2-binding protein 1 (TopBP1), which interacts with the C-terminal domain of Rad9 and activates ATR–ATRIP (Figure 4B) [81–83]. This set of reactions ultimately leads to cell cycle arrest and ensures the conclusion of DNA synthesis.

Trypanosomatids dwell in inhospitable environments in which replication stress may arise and cause a threat to their genome's stability. *Leishmania* spp. replicate under conditions of high oxidative stress within host macrophages [54,84], *T. brucei* proliferates in the host bloodstream [85], and *T. cruzi* multiplication is inhibited for long periods in chronic infections [86]. The genome architecture and transcription mechanism of trypanosomatids may favor replication stress: because virtually all genes are cotranscribed from multigene transcription units that may include hundreds of genes [2,87], the long-distance and near constitutive movement of RNA Pol II across the genome must increase the severity of replication–transcription conflicts [88].

The trypanosomatid genome encodes many homologous proteins of the replication stress response, and their functional characterization, which has only begun, has already revealed remarkable features. As mentioned before, trypanosomatid RPA-1 lacks the 70N domain (Box 2), which in other eukaryotes is required for RPA binding to the 9-1-1 clamp and ATR activation [89]. In *Leishmania*, RPA-1 associates with chromatin in response to replication stress and colocalizes with Hus1 and Rad9 [90]. It is possible that the interplay between RPA, 9-1-1, and ATR is mediated by another still unknown protein in this parasite, which could not only supply the absent 70N domain but also serve as a distinct point of regulation for the replication stress response.

Homologs of all the 9-1-1 subunits have been identified in trypanosomatids. In *Leishmania*, all three subunits associate with chromatin in response to replication stress, and Rad9- or Hus1-deficiency impacts replication stress and DSB responses [90]. Primary sequence and structural predictions indicate that the degree of conservation is not the same for the three subunits, suggesting the existence of a distinct selective pressure that drives the evolution of these subunits in *Leishmania*. Rad9 is the most divergent subunit with an expanded C-terminal domain [91], which harbors key phosphorylation sites required for the genotoxic stress response in other eukaryotes [92,93]. Such structural divergence of Rad9 suggests that the functional regulation of 9-1-1 might also present relevant peculiarities. Mammals and yeast have evolved paralogs and isoforms of 9-1-1 subunits, which possibly allow them to engage in compartmentalized functions [94,95]. In *Leishmania*, the apparent absence of 9-1-1 paralogs has possibly been compensated for by the formation of alternative complexes [90]. Two sets of evidence corroborate this idea. First, Rad9 exists in a complex distinct from 9-1-1, and Hus1 is found in a monomeric form. Second, the phenotypes observed in response to replication stress and DSB formation in Rad9-deficient cells are distinct from those associated with Hus1 deficiency. Hence, functional flexibility of the 9-1-1 complex in *Leishmania* is probably enabled by the association of the subunits in different complexes within the cell.

Currently, it is unknown whether ATR is required for the replication stress response of trypanosomatids, but kinase activity is necessary for the replication stress response in *Leishmania*. For instance, phosphorylation of H2A is triggered by replication stress in a Rad9- and Hus1-dependent manner [90], but the specific kinase activity linked to this process has not been identified [96]. Another common response to replication stress in eukaryotes is activation of dormant origins [97,98]. A recent study found evidence that replication stress can trigger

activation of at least one putative dormant origin in *T. brucei* [12], but this has not been expanded to a genome-wide scale or to other trypanosomatids. Furthermore, no *T. brucei* proteins or pathways that act in dormant origin activation have been identified and, as we have noted above, we cannot rule out origin-independent initiation of replication after stress, perhaps in *Leishmania* or *T. cruzi* particularly.

### Concluding Remarks

There are many differences emerging in the structural composition of the protein complexes involved in DNA maintenance between trypanosomatids and eukaryotic models, though in most cases the advantages and disadvantages of the presence or absence of particular domains and/or motifs in specific proteins remains unknown (see Outstanding Questions). What was previously thought to be a single-component ORC complex (ORC1/CDC6), is now likely a multimeric complex. In addition, further studies are necessary to reveal the complete composition of the pre-RC, as well as the action of ORC1B. There is insufficient evidence to infer the mechanism used by trypanosomatids to prevent re-replication, although clues point to CDC45 as a possible inhibitor of replication outside the S phase [22]. On the other hand, there is evidence pointing to pre-RC components as key factors in blocking DNA replication in infective forms of *T. cruzi* [27]. In addition, we have only started unveiling the replication stress response in these parasites, and the establishment of a unique response to this phenomenon may have shaped some of the peculiar genome phenomena observed in these parasites, including the remarkable expression of VSG genes in *T. brucei* and the extraordinary genome plasticity found in *Leishmania* species. Further understanding of the pre-RC and 9-1-1 components, as well as identification of pathways implicated in the response to replication stress, may provide a basis for the design of more effective chemotherapy arsenals against the devastating diseases caused by trypanosomatids.

### Acknowledgments

The authors are grateful to the São Paulo Research Foundation (FAPESP) – Center of Toxins, Immune Response and Cell Signaling (CeTICS): grants 2014/24170-5, 2013/07467-1, 2015/10580-0, and 2016/50050-2; to the National Council for Scientific and Technological Development (CNPq): grant 304329/2015-0; to the TriTrypDB: the kinetoplastid genomics resource; and to [timetree.org](http://timetree.org). Work in RM's laboratory that contributed to this article was supported by the Wellcome Trust [089172, 083485], the BBSRC [BB/K006495/1, BB/N016165/1], RCUK-CONFAP [BB/M028909/1], Fundação para a Ciência e Tecnologia [SFRH/BD/68784/2010], and the European Commission [RECREPEMLE]. The Wellcome Centre for Molecular Parasitology is supported by core funding from the Wellcome Trust [104111], and next-generation sequencing was conducted in Glasgow Polyomics, which is also supported by the Wellcome Trust [105614].

### References

- da Silva, M.S. *et al.* (2017) Differences in the detection of BrdU/EdU incorporation assays alter the calculation for G1, S, and G2 phases of the cell cycle in trypanosomatids. *J. Eukaryot. Microbiol.* Published online March 4, 2017. <http://dx.doi.org/10.1111/jeu.12408>
- El-Sayed, N.M. *et al.* (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309, 404–409
- Simpson, A. *et al.* (2004) Early evolution within kinetoplastids (Euglenozoa), and the late emergence of trypanosomatids. *Protist* 155, 407–422
- Flegontov, P. *et al.* (2013) *Paratrypanosoma* is a novel early-branching trypanosomatid. *Curr. Biol.* 23, 1787–1793
- Simpson, A.G.B. *et al.* (2006) The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol.* 22, 168–174
- Oppeides, F.R. *et al.* (2016) Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *J. Eukaryot. Microbiol.* 63, 657–678
- Albalat, R. and Cañestro, C. (2016) Evolution by gene loss. *Nat. Rev. Genet.* 17, 379–391
- Genois, M.M. *et al.* (2014) DNA repair pathways in trypanosomatids: from DNA repair to drug resistance. *Microbiol. Mol. Biol. Rev.* 78, 40–73
- Laffitte, M.N. *et al.* (2016) Plasticity of the *Leishmania* genome leading to gene copy number variations and drug resistance. *F1000Research* 5, 2350
- Tiengwe, C. *et al.* (2012) Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. *Cell Rep.* 2, 185–197
- Marques, C.A. *et al.* (2015) Genome-wide mapping reveals single-origin chromosome replication in *Leishmania*, a eukaryotic microbe. *Genome Biol.* 16, 230
- Calderano, S.G. *et al.* (2015) Single molecule analysis of *Trypanosoma brucei* DNA replication dynamics. *Nucleic Acids Res.* 43, 2655–2665
- Stanojic, S. *et al.* (2016) Single-molecule analysis of DNA replication reveals novel features in the divergent eukaryotes *Leishmania* and *Trypanosoma brucei* versus mammalian cells. *Sci. Rep.* 6, 23142

### Outstanding Questions

Are there functional homologs for CDC6 and Cdt1 in trypanosomatids?

Could the ORC complex recruit MCM<sub>2-7</sub> and bind to DNA without ORC1B?

Is there a need for kinase-mediated regulation of the trypanosomatid pre-RC if ORC1B provides a positive, activating function?

If ORC remains bound to the origins after replisome recruitment, how does the replisome bypass it?

Why is the DNA replication rate in trypanosomatids slightly higher compared with model eukaryotes?

What is the mechanism used by these protozoans to avoid DNA re-replication during and after the S phase?

Are there other proteins involved in the recruitment of the 9-1-1 clamp and activation of ATR in response to replication stress?

What are the advantages or disadvantages of the divergent structures of some replication machinery proteins?

What mechanisms are used by trypanosomatids to resolve the predicted severe conflicts between transcription and replication that arise due to multi-genic transcription?

Does the execution and/or control of genome replication lead to genome plasticity in trypanosomatids?

Do trypanosomatids use widespread origin-independent replication initiation to complete synthesis of their genomes?

14. Lombraña, R. *et al.* (2016) Transcriptionally driven DNA replication program of the human parasite *Leishmania major*. *Cell Rep.* 16, 1774–1786
15. Rocha-Granados, M.C. and Klingbeil, M.M. (2016) *Leishmania* DNA replication timing: a stochastic event? *Trends Parasitol.* 32, 755–757
16. Bleichert, F. *et al.* (2015) Crystal structure of the eukaryotic origin recognition complex. *Nature* 519, 321–326
17. Tocilj, A. *et al.* (2017) Structure of the active form of human origin recognition complex and its ATPase motor module. *Elife* 6, e20818
18. Fernández-Cid, A. *et al.* (2013) An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamers assembly. *Mol. Cell* 50, 577–588
19. Frigola, J. *et al.* (2013) ATPase-dependent quality control of DNA replication origin licensing. *Nature* 495, 339–343
20. Tica, S. *et al.* (2015) Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* 161, 513–525
21. Tiengwe, C. *et al.* (2012) Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture. *PLoS One* 7, e32674
22. Dang, H.Q. and Li, Z. (2011) The Cdc45-Mcm2-7-GINS protein complex in trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in the origin recognition complex. *J. Biol. Chem.* 286, 32424–32435
23. Godoy, P.M. *et al.* (2009) Trypanosome prereplication machinery contains a single functional ORC1/CDC6 protein, which is typical of archaea. *Eukaryot. Cell* 8, 1592–1603
24. Marques, C.A. *et al.* (2016) Diverged composition and regulation of the *Trypanosoma brucei* origin recognition complex that mediates DNA replication initiation. *Nucleic Acids Res.* 44, 4763–4784
25. Benmerzouga, I. *et al.* (2013) *Trypanosoma brucei* Orc1 is essential for nuclear DNA replication and affects both VSG silencing and VSG switching. *Mol. Microbiol.* 87, 196–210
26. Sheader, K. *et al.* (2004) Bloodstream form-specific up-regulation of silent vsg expression sites and procyclin in *Trypanosoma brucei* after inhibition of DNA synthesis or DNA damage. *J. Biol. Chem.* 279, 13363–13374
27. Calderano, S. *et al.* (2014) ORC1/CDC6 and MCM7 distinct associate with chromatin through *Trypanosoma cruzi* life cycle. *Mol. Biochem. Parasitol.* 193, 110–113
28. Fernández-Orgil, A. *et al.* (2016) A putative *Leishmania* DNA polymerase theta protects the parasite against oxidative damage. *Nucleic Acids Res.* 44, 4855–4870
29. Pavani, R.S. *et al.* (2016) Replication protein A presents canonical functions and is also involved in the differentiation capacity of *Trypanosoma cruzi*. *PLoS Negl. Trop. Dis.* 10, e0005181
30. Rudd, S.G. *et al.* (2013) PPL2 translesion polymerase is essential for the completion of chromosomal DNA replication in the African trypanosome. *Mol. Cell* 52, 554–565
31. Leman, A.R. and Noguchi, E. (2013) The replication fork: understanding the eukaryotic replication machinery and the challenges to genome duplication. *Genes (Basel)* 4, 1–32
32. MacNeill, S. (ed.) (2012) *The Eukaryotic Replisome: A Guide to Protein Structure and Function*, p. 62, Springer
33. Silva, M.S. and Cano, M.I.N., eds (2017) *Molecular and Cellular Biology of Pathogenic Trypanosomatids*, Bentham Science Publishers
34. Kumar, D. *et al.* (2009) The distribution pattern of proliferating cell nuclear antigen in the nuclei of *Leishmania donovani*. *Microbiology* 155, 3748–3757
35. Calderano, S.G. *et al.* (2011) *Trypanosoma cruzi* DNA replication includes the sequential recruitment of pre-replication and replication machineries close to nuclear periphery. *Nucleus* 2, 136–145
36. Elias, M.C.E. *et al.* (2002) Chromosome localization changes in the *Trypanosoma cruzi* nucleus. *Eukaryot. Cell* 1, 944–953
37. Kaufmann, D. *et al.* (2012) Regulation and spatial organization of PCNA in *Trypanosoma brucei*. *Biochem. Biophys. Res. Commun.* 419, 698–702
38. Langston, L.D. and O'Donnell, M. (2008) DNA polymerase delta is highly processive with proliferating cell nuclear antigen and undergoes collision release upon completing DNA. *J. Biol. Chem.* 283, 29522–29531
39. Jones, R.M. and Petermann, E. (2012) Replication fork dynamics and the DNA damage response. *Biochem. J.* 443, 13–26
40. Ponce, I. *et al.* (2017) A flap endonuclease (TcFEN1) is involved in *Trypanosoma cruzi* cell proliferation, DNA repair, and parasite survival. *J. Cell Biochem.* 118, 1722–1732
41. Zou, L. and Stillman, B. (2000) Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbp4p kinase. *Mol. Cell Biol.* 20, 3086–3096
42. Liu, T. and Huang, J. (2016) Replication protein A and more: single-stranded DNA-binding proteins in eukaryotic cells. *Acta Biochim. Biophys. Sin. (Shanghai)* 48, 665–670
43. Braun, K.A. *et al.* (1997) Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase alpha by multiple mechanisms. *Biochemistry* 36, 8443–8454
44. da Silveira, R.C.V. *et al.* (2013) The natural absence of RPA1N domain did not impair *Leishmania amazonensis* RPA-1 participation in DNA damage response and telomere protection. *Parasitology* 140, 547–559
45. Pavani, R.S. *et al.* (2014) RPA-1 from *Leishmania amazonensis* (LaRPA-1) structurally differs from other eukaryote RPA-1 and interacts with telomeric DNA via its N-terminal OB-fold domain. *FEBS Lett.* 588, 4740–4748
46. Brown, G.W. *et al.* (1994) Isolation of the genes encoding the 51-kilodalton and 28-kilodalton subunits of *Crithidia fasciculata* replication protein A. *Mol. Biochem. Parasitol.* 63, 135–142
47. Brown, L.M. and Ray, D.S. (1997) Cell cycle regulation of RPA1 transcript levels in the trypanosomatid *Crithidia fasciculata*. *Nucleic Acids Res.* 25, 3281–3289
48. Pommier, Y. *et al.* (2016) Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* 17, 703–721
49. Fragoso, S.P. *et al.* (1998) Expression and cellular localization of *Trypanosoma cruzi* type II DNA topoisomerase. *Mol. Biochem. Parasitol.* 94, 197–204
50. Kulikowicz, T. and Shapiro, T.A. (2006) Distinct genes encode type II Topoisomerases for the nucleus and mitochondrion in the protozoan parasite *Trypanosoma brucei*. *J. Biol. Chem.* 281, 3048–3056
51. Das, B.B. *et al.* (2006) Topoisomerase research of kinetoplastid parasite *Leishmania*, with special reference to development of therapeutics. *Indian J. Med. Res.* 123, 221–232
52. Wu, P. *et al.* (2012) Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. *Cell* 150, 39–52
53. Soudet, J. *et al.* (2014) Elucidation of the DNA end-replication problem in *Saccharomyces cerevisiae*. *Mol. Cell* 53, 954–964
54. da Silva, M.S. *et al.* (2017) Consequences of acute oxidative stress in *Leishmania amazonensis*: from telomere shortening to the selection of the fittest parasites. *Biochim. Biophys. Acta – Mol. Cell Res.* 1864, 138–150
55. Lira, C.B.B. *et al.* (2007) Telomere biology of trypanosomatids: beginning to answer some questions. *Trends Parasitol.* 23, 357–362
56. Sekedat, M.D. *et al.* (2010) GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol. Syst. Biol.* 6, 353
57. Conti, C. *et al.* (2007) Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. *Mol. Biol. Cell* 18, 3059–3067
58. Yeeles, J.T.P. *et al.* (2017) How the eukaryotic replisome achieves rapid and efficient DNA replication. *Mol. Cell* 65, 105–116
59. Maree, J.P. and Patterson, H.G. (2014) The epigenome of *Trypanosoma brucei*: a regulatory interface to an unconventional transcriptional machine. *Biochim. Biophys. Acta* 1839, 743–750

60. Duraisingh, M.T. and Horn, D. (2016) Epigenetic regulation of virulence gene expression in parasitic protozoa. *Cell Host Microbe* 19, 629–640
61. Elsasser, S. *et al.* (1999) Phosphorylation controls timing of Cdc6p destruction: a biochemical analysis. *Mol. Biol. Cell* 10, 3263–3277
62. Drury, L.S. *et al.* (2000) The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Curr. Biol.* 10, 231–240
63. Lygerou, Z. *et al.* (2000) The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* 404, 625–628
64. Nguyen, V.Q. *et al.* (2000) Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2–7. *Curr. Biol.* 10, 195–205
65. Nguyen, V.Q. *et al.* (2001) Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* 411, 1068–1073
66. Liku, M.E. *et al.* (2005) CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2–7 complex prevents chromosomal rereplication. *Mol. Biol. Cell* 16, 5026–5039
67. Chen, S. and Bell, S.P. (2011) CDK prevents Mcm2–7 helicase loading by inhibiting Cdt1 interaction with Orc6. *Genes Dev.* 25, 363–372
68. Wei, L. and Zhao, X. (2016) A new MCM modification cycle regulates DNA replication initiation. *Nat. Struct. Mol. Biol.* 23, 209–216
69. Parker, M.W. *et al.* (2017) Mechanisms and regulation of DNA replication initiation in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* 52, 107–144
70. Findeisen, M. *et al.* (1999) Cyclin A-dependent kinase activity affects chromatin binding of ORC, Cdc6, and MCM in egg extracts of *Xenopus laevis*. *Eur. J. Biochem.* 264, 415–426
71. Blow, J.J. and Dutta, A. (2005) Preventing re-replication of chromosomal DNA. *Nat. Rev. Mol. Cell Biol.* 6, 476–486
72. Kumar, D. *et al.* (2008) Expression and subcellular localization of ORC1 in *Leishmania major*. *Biochem. Biophys. Res. Commun.* 375, 74–79
73. Alexander, J.L. *et al.* (2015) Replication fork progression during re-replication requires the DNA damage checkpoint and double-strand break repair. *Curr. Biol.* 25, 1654–1660
74. Reis-Cunha, J.L. *et al.* (2015) Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct *Trypanosoma cruzi* strains. *BMC Genomics* 16, 499
75. Callejas, S. *et al.* (2006) Hemizygous subtelomeres of an African trypanosome chromosome may account for over 75% of chromosome length. *Genome Res.* 16, 1109–1118
76. Fanning, E. *et al.* (2006) A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res.* 34, 4126–4137
77. Zou, L. and Elledge, S.J. (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* 300, 1542–1548
78. Syljuasen, R.G. *et al.* (2005) Inhibition of human Chk1 causes increased initiation of DNA replication, phosphorylation of ATR targets, and DNA breakage. *Mol. Cell. Biol.* 25, 3553–3562
79. Zhao, H. and Piwnicka-Worms, H. (2001) ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. *Mol. Cell. Biol.* 21, 4129–4139
80. Ellison, V. and Stillman, B. (2003) Biochemical characterization of DNA damage checkpoint complexes: clamp loader and clamp complexes with specificity for 5' recessed DNA. *PLoS Biol.* 1, e33
81. Delacroix, S. *et al.* (2007) The Rad9-Hus1-Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. *Genes Dev.* 21, 1472–1477
82. Lee, J. and Dunphy, W.G. (2010) Rad17 plays a central role in establishment of the interaction between TopBP1 and the Rad9-Hus1-Rad1 complex at stalled replication forks. *Mol. Biol. Cell* 21, 926–935
83. Lee, J. *et al.* (2007) The Rad9-Hus1-Rad1 checkpoint clamp regulates interaction of TopBP1 with ATR. *J. Biol. Chem.* 282, 28036–28044
84. Naderer, T. and McConville, M.J. (2011) Intracellular growth and pathogenesis of *Leishmania* parasites. *Essays Biochem.* 51, 81–95
85. Pays, E. *et al.* (2014) The molecular arms race between African trypanosomes and humans. *Nat. Rev. Microbiol.* 12, 575–584
86. Silva-dos-Santos, D. *et al.* (2017) Unraveling Chagas disease transmission through the oral route: gateways to *Trypanosoma cruzi* infection and target tissues. *PLoS Negl. Trop. Dis.* 11, e0005507
87. Reynolds, D. *et al.* (2014) Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in *Leishmania major* and *Trypanosoma brucei*. *Nucleic Acids Res.* 42, 9717–9729
88. Helmrich, A. *et al.* (2013) Transcription-replication encounters, consequences and genomic instability. *Nat. Struct. Mol. Biol.* 20, 412–418
89. Xu, X. *et al.* (2008) The basic cleft of RPA70N binds multiple checkpoint proteins, including RAD9, to regulate ATR signaling. *Mol. Cell. Biol.* 28, 7345–7353
90. Damasceno, J.D. *et al.* (2016) Functional compartmentalization of Rad9 and Hus1 reveals diverse assembly of the 9-1-1 complex components during the DNA damage response in *Leishmania*. *Mol. Microbiol.* 101, 1054–1068
91. Damasceno, J.D. *et al.* (2013) LmHus1 is required for the DNA damage response in *Leishmania major* and forms a complex with an unusual Rad9 homologue. *Mol. Microbiol.* 90, 1074–1087
92. Furuya, K. *et al.* (2010) DDK phosphorylates checkpoint clamp component Rad9 and promotes its release from damaged chromatin. *Mol. Cell* 40, 606–618
93. Navadgi-Patil, V.M. and Burgers, P.M. (2009) A tale of two tails: activation of DNA damage checkpoint kinase Mec1/ATR by the 9-1-1 clamp and by Dpb11/TopBP1. *DNA Repair (Amst)* 8, 996–1003
94. Janes, S. *et al.* (2012) Heat induction of a novel Rad9 variant from a cryptic translation initiation site reduces mitotic commitment. *J. Cell Sci.* 125, 4487–4497
95. Pérez-Castro, A.J. and Freire, R. (2012) Rad9B responds to nucleolar stress through ATR and JNK signalling, and delays the G1-S transition. *J. Cell Sci.* 125, 1152–1164
96. Glover, L. and Horn, D. (2012) Trypanosomal histone  $\gamma$ H2A and the DNA damage response. *Mol. Biochem. Parasitol.* 183, 78–83
97. Chen, Y.H. *et al.* (2015) ATR-mediated phosphorylation of FANCI regulates dormant origin firing in response to replication stress. *Mol. Cell* 58, 323–338
98. Toledo, L.I. *et al.* (2013) ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell* 155, 1088–1103
99. Tandon, R. *et al.* (2014) Characterization of the proliferating cell nuclear antigen of *Leishmania donovani* clinical isolates and its association with antimony resistance. *Antimicrob. Agents Chemother.* 58, 2997–3007